Mem. S.A.It. Suppl. Vol. 13, 147

Memorie della
Supplementi

Making the Grid and the Virtual Observatory interoperable

G. Taffoni, C. Vuerli, and F. Pasian

Istituto Nazionale di Astrofisica – Unità Sistemi Informativi and Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy, e-mail: taffoni@oats.inaf.it

Abstract. The term "grid" in the Virtual Observatory context has mainly been used to indicate a set of interoperable services. This is rather different from the approach other scientific communities are taking mainly based on using the grid for computational tasks. Within this framework, it appears as extremely important to interconnect the Virtual Observatory and the computational grid infrastructures. Harmonisation of the Virtual Observatory infrastructure and user tools with the developments being carried out within the various national and European grid projects is an important goal to achieve.
We present the point of view of followed in the framework of the EuroVO Data Centre Alliance project, that we will propose to the International Virtual Observatory Alliance as a successful example of interoperability.

Key words. Stars: abundances – Grid: interoperability – Grid: data management – Galaxy: single-sign-on – Virtual Observatory: interoperability – Virtual Observatory: EuroVO

1. Introduction

The dramatic improvement of Internet facilities and technologies together with the improvement of hardware devices both in capacity and capability drove the evolution of a new idea of distributed computing: the Grid concept.

Although the motivations and goals for grids are obvious, there is no clear definition for a grid system. The grid is a framework for "flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources" (Foster, Kesselman & Tuecke 2001), "a single seamless computational environment in which cycles, communication, and data are shared, and in which the workstation across the continent is no less than one down the hall" (Grimshaw et al. 1994), "a widearea environment that transparently consists of workstations, personal computers, geographic rendering engines, supercomputers, and nontraditional devices: e.g., TVs, toasters, etc." (Grimshaw & Wulf 1996), "a collection of geographically separated resources (people, computers, instruments, databases) connected by a high speed network [...distinguished by...] a software layer, often called middleware, which transforms a collection of independent resources into a single, coherent, virtual machine" (Lindahl et al. 1998). With varying degrees of precision these definitions describe the central notion of grid systems. More generally we can say that we have a grid any time a scientific community has a commodity that want to share and develops a discovery system to locate and access it.

In this paper we present the work done in the framework of the EuroVO Data Center Alliance project to allow the Data Centers to take full advantage of the grid computational and storage resources. And we discuss the interoperability issues of the International Virtual Observatory (IVO) and computational and data grids, in particular the EGEE grid (Laure et al. 2006).

2. The international Virtual Observatory

Large digital sky surveys and data archives are becoming the principal sources of data in astronomy. Combining surveys done at different wavelengths, from radio and infrared, through visible light, ultraviolet, and x-rays, both from the ground-based telescopes and from space observatories, provide a new, panchromatic picture of our Universe, and lead to a better understanding of the objects in it. To profit of all those data astronomers need to face a great technological challenge: these vast amounts of information must be managed, combined, analyzed and explored in a quick and efficient way. The concept of a virtual observatory (VObs) thus emerged. A VObs would be a set of federated, geographically distributed, major digital sky archives, with the software tools and infrastructure to combine them in an efficient and user-friendly manner, and to explore the resulting data sets whose sheer size and complexity are beyond the reach of traditional approaches. The International Virtual Observatory Alliance (IVOA) was formed in 2002 with the mission to facilitate the international coordination and collaboration necessary to develop and deploy tools, systems and organizational structures necessary to enable the international utilization of heterogeneous astronomical data sources as an integrated and interoperating VObs (Solano 2006).

3. The EuroVO Data Centre Alliance

EuroVO is a specifically European implementation of the VObs idea that will produce a unified virtual data and service resource with the ability to perform complex data discovery and manipulation tasks across the whole range of astronomical research topics. The EuroVO Data Centre Alliance (http://www.euro-vo.org/pub/dca/) helps and encourages European data-centres to publish data in the VObs.

One of the key aspects of the Data Centre Alliance activity regards the access to massive computing facilities that are particularly relevant for the theoretical astronomy but necessary also for the generic Astronomical Data Centre. There is significant investment in grid computing at national and European level from outside astronomy. The results of this investment are powerful grid infrastructures at disposal of European astronomers. The Work Package 5 (WP5) of EuroVO-DCA aims at achieving coordination between the VObs and the computational grid communities. It organizes the activity necessary to allow Virtual Observatory users to exploit (through the data centres involved in the project) the processing and storage capabilities offered by the computational grid projects. WP5 aims to make the grid infrastructures useful and usable to users of the VObs (Taffoni, Maino et al. 2007).

WP5 collates information on de-jure and de-facto standards of the grid in order to make it easier for data centres to learn these standards.

4. The Virtual Observatory Grid

The key idea that drove the development and deployment of the IVO infrastructure is the possibility to locate and share Astronomical data. The VObs is actually a (database structured) data and services grid. Making the VObs interoperable with the computational Grids implies to identify some key standards to adopt and to design a number of adaptors (interfaces) that can be used to access the grid resources. The common standards regards the authentication and authorization procedures and a common protocol to access data.

Grid services (ex. data management and job management services) have restricted access. The access to the service is secured by the

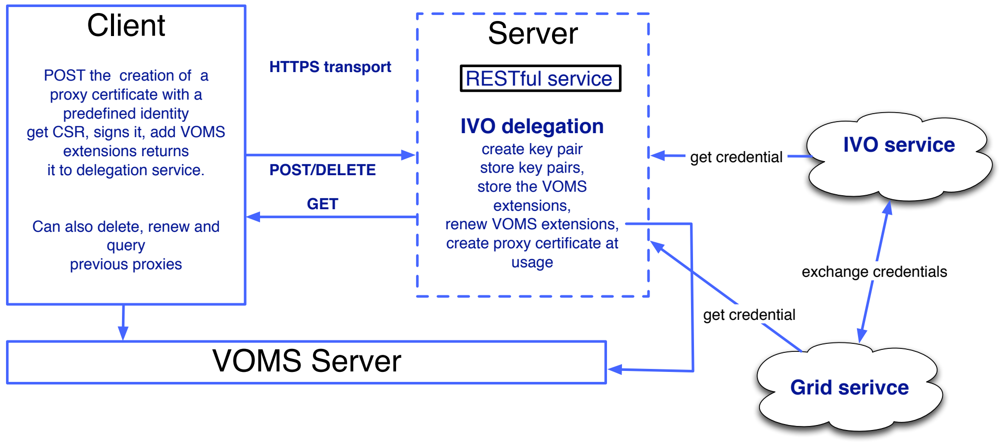

Fig. 1. A schematic view of the VObs delegation service with full support of the VOMS authorization system.

grid Security Infrastructure (GSI) mechanisms (Foster & Kesselman 1997). GSI uses public key cryptography (also known as asymmetric cryptography) as the basis for its functionality. A central concept in GSI authentication is the certificate. Every user and service on the grid is identified via a certificate, which contains information vital to identifying and authenticating the user or service. When a client program makes a requests to one of these secured services, it is typically doing so in the name of the user running the client. I.e., the client presents to the service credentials authenticating the user's identity.

GRIDFTP is the protocol used by grid systems to move data across wide area networks, it is already widely used as a common protocol to make grid infrastructure to interoperate (e.g. EGEE, NAREGI, OSG). Moreover, it is possible to access FITS file using GRIDFTP protocol from grid storage systems. A GRIDFTP driver has been implemented in CFITSIO (Taffoni, Barisani et al. 2007).

The adaptors are the interfaces to the specific grid environment. For example to access EGEE grid it is necessary to join an EGEE Virtual Organization and to use the EGEE authorization system (the Virtual Organization Membership Service - VOMS). The adaptors should be designed as plug-ins of the standard IVO services.

The final goal of this approach is to give the IVO users the possibility to use the computational and data grid infrastructure. The entry point to the grids is the VObs itself that (if necessary) hide all the complexity of the grid systems. Access to the grid is mediated by IVO protocols and standards.

5. An example of VObs and GRID interoperable service: credential delegation and VOSpace

When a user makes use of a remote resource (e.g. a VObs resource that querying data from a database) that resource may in turn need to access third-party resources (e.g. data repositories in the grid environment) on behalf of the user in order to complete the task. Such access may possibly span across multiple security domains. In this scenario, it is necessary to delegate (parts of) the users rights to the remote resource such that it in turn can access the necessary third parties.

The credential-delegation protocol allows a client programme to delegate a user's credentials to a service, that service then makes requests of other services in the name of that user. The delegation protocol defines a

REST web-service for delegation of identity credentials between VObs installations Plante, Rixon, Taffoni (2007). At present, the IVOA delegation profile is in draft state and the Delegation Web Service implementation is in testing phase.

EGEE-II uses VOMS to express authorization. In order to run a grid ”job”, a VObs service must delegate a credential to the grid and also direct the grid service to a VOMS installation that expresses the user’s right to use the grid. We extend the capability of the VObs delegation service in order to support VOMS extensions and to manage them (contacting the VOMS server to initialize on-demand the extensions, renewing the extension when necessary or verify the credentials).

This modified version of the protocol is a first step towards interoperability. In that way a client can use both EGEE-II grid and VObs resources and make them interact (Fig. 1).

VOSpace is the virtual filesystem for IVO. It provides a protocol and the APIs to interface over a disk space, a distributed filesystem as the Storage Resource Broker, or even a grid file system. VOSpace is a thin software layer that can be used to virtualize different data resources. We design a VOSpace ”plug-in” that can be used to access the EGEE grid filesystem using the IVO standards.

6. Conclusions

It is important to make the VObs grid to interoperate with the computational and data grid resources. This will produce a ”cyber - infrastructure” that will support Astronomers in any aspect of their research activity, from data discovery and query to computation, from data storage to sharing resources and files. This infrastructure will foster the collaboration and increase the scientific production.

The delegation service is one step towards the interoperability, work is going on for what regards data management and job management.

Acknowledgements. This work has been done in the framework of the EuroVO Data Centre Alliance, funded by the EU within it 6^{th} Framwork Program (FP6).

References

Foster, I., Kesselman C., & Tuecke S. (2001). The Int J. HPC Appl 15: 200

Foster, I. & Kesselman C. (1997). The Int J. Supercomput. Appl, 11: 115

Grimshaw A. S. & Wulf W. (1996), HPDC 1996: 89-99

Grimshaw A. S. et al. (1994), Tech Rep No. CS-94-2, June 1994

Laure, E., Fisher, S.M., Frohner, A., et al. (2006) Comp. Meth. in Science and Technology 12(1), 33-45

Lindahl G., Grimshaw A., Ferrari A., Holcomb K. (1998), Tech Rep No. white paper

Plante R., Rixon G., Taffoni G., 2007 IVOA credential delegation protocol, IVOA WD

Solano, E.. (2006). Astronomical Society of the Pacific Conference Series: 29

Taffoni G., Maino D., Vuerli C., Castelli G., Smareglia R., Zacchei A., Pasian F., (2007). Future Generation Comp. Syst. 23(2): 189-200

Taffoni, G., Barisani, A., Vuerli, C., Pence, W., Pasian, F. (2007), ASPC, 376, 89